DOCUMENT RESUME

ED 449 169 TM 032 247

AUTHOR	Tanguma, Jesus
TITLE	Equating Test Scores Using the Linear Method: A Primer.
PUB DATE	2000-01-27
NOTE	18p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 27-29, 2000).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Equated Scores; *Test Format
IDENTIFIERS	*Anchor Tests; *Linear Equating Method

ABSTRACT

This paper describes four commonly used designs in equating test scores. These designs are: (1) single-group; (2) random-group; (3) equivalent-group; and (4) anchor-test. Each design requires that its data be collected according to specific guidelines. Three of the four methods are illustrated through hypothetical examples. All four methods try to equate test scores from equally reliable and parallel measures. Although the anchor-test design is not as simple to implement as the other designs, it is one of the most popular equating procedures. (Contains 2 tables, 1 figure, and 10 references.) (Author/SLD)

Running head: LINEAR EQUATING

Equating Tests Scores Using the Linear Method: A Primer

Jesus Tanguma

Texas A & M University

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, January 27, 2000.

Abstract

The present paper presents four commonly used designs in equating test scores. These designs are (a) single-group, (b) random-group, (c) equivalent-group, and (d) anchor-test. For each design, its data must be collected according to specific guidelines. Three of the four methods are illustrated by means of hypothetical situations. All four methods try to equate test scores from equally reliable and parallel measures. Although the anchor-test design is not as simple to implement as the other designs, it is one of the most popular equating procedures

Equating Tests Scores Using the Linear Method: A Primer

Suppose student A is administered the fall edition of test X and gets a grade, GAX, on it. Also, suppose that student B is administered the spring edition of test X and gets a grade, GBX, on it. Can the two scores be compared given that they come from different test administrations? To make matters even worse, assume that student B is taking test X for the second time. Can it be safely said that student B has no advantage over student A due to the fact that he/she is not only taking a later administration of the test but also re-testing? It is for these reasons, among others, that many testing programs use multiple editions of a given test. That is, most testing programs use a new set of questions, as similar in difficulty and content as possible, on each test administration.

Since every test administration uses different editions of the test and different editions of the test use different sets of questions, it follows that there will be some differences among test editions. In other words, although test developers attempt to develop test editions that are as similar as possible in content and statistical specifications, there will be some differences in the level of difficulty. Consequently, if there is to be any comparison among examinees who have taken different editions of the same test, a process that would produce comparable scores on these editions must be sought. One such process is to equate the test scores.

The purpose of this paper is to present a brief introduction to the Linear Equating Method. According to Kolen (1988),

> In linear equating the means and standard deviations on the two forms for a particular group of examinees are set equal. In this method, Form 2 scores are converted so as to have the same mean and standard deviation as scores on Form 1. (p. 33)

Four commonly used methods will be discussed, but only three will be illustrated by means of hypothetical situations.

Definition of Equating

A review of the literature offers several definitions of equating. Angoff (1982), points out that

> Equating is the process of developing a conversion from the system of units of one form of a test to the system of units of another form so that scores derived from the two forms after conversion will be equivalent and interchangeable. (p. 56)

Therefore, when equating has been properly done, "it is possible to compare directly the performances of two individuals who have taken different forms of a test" (Angoff, 1971, p. 563).

According to Lord (1980), two tests forms, X and Y, are considered equated if it is a matter of indifference to each examinee which test he or she takes. Moreover, Petersen, N. S., Kolen, M. J., and Hoover, H. D. (1986) have pointed out that scores on tests X and Y are considered equated only when the following conditions hold:

1. Same Ability-the two tests must both be measures of the same characteristics (latent trait, ability, or skill).
2. Equity- for every group of examinees of identical ability, the conditional frequency distribution of scores on test X, after transformation, is the same as the conditional frequency distribution on test Y.
3. Population Invariance-the transformation is the same regardless

of the group from which it is derived.

4. Symmetry-the transformation is invertible, that is the mapping of scores from Form X to From Y is the same as the mapping of scores from Form Y to Form X.

The equity requirement follows from Lord's (1980) statement and has the following implications:

1. Tests measuring different traits or abilities cannot be equated.
2. Raw scores on unequally reliable tests cannot be equated (since otherwise scores from unreliable test can be equated to scores on a reliable test, thereby obviating the need for constructing reliable test!).
3. Raw scores on tests with varying difficulty levels, i.e., in vertical equating situations, cannot be equated (since in this case the true scores will have a nonlinear relation and the tests therefore will not be equally reliable at different ability levels).
4. The conditional frequency distribution at ability level θ, $f[x:\theta]$ of score θ on test X is the same as the conditional frequency distribution for the transformed score $x(y)$, $f[x(y):\theta]$, where $x(y)$ is a one-to-one function of y.
5. Fallible scores on tests X and Y cannot be equated unless tests X and Y are strictly parallel (since the condition of identical conditional frequency distributions, under regularity conditions, implies that the moments of the two distributions are equal).

6. Perfectly reliable tests can be equated.

Designs for Equating

Before an equating may be done, data must be collected from very specific designs. Four commonly used designs to collect data before performing an equating are (a) single-group design; (b) random-groups design; (c) equivalent-group design; and (d) anchor-test design, see Figure 1. When using the single-group design, both tests are administered to the same group of examinees. The forms are administered one after the other, and when possible on the same day. As Crocker and Algina (1986) point out, "since the same examinees take both tests, the difficulty levels of the tests are not confounded with the ability levels of the examinees" (p. 198). However, in doing so the researcher must assume that there is no practice and/or fatigue effect on the scores on the second test (Petersen et al., 1986, p. 245).

Insert Figure 1 About Here

A common solution to the above mentioned problem is to package the books in such a way that every other booklet is a new form. This way, every other examinee takes a new form. According to Angoff (1971) "this procedure will fail to yield randomly equivalent groups only when the examinees themselves are seated in a sequence (e.g. boy, girl, boy, girl, etc.) that may be correlated with the test score" (p. 569).

When the equating is being done by means of the equivalent-group design, the two forms are administered to two random groups, one for each group. This way, every examinee takes only one test. However, since every examinee takes only one

test, there is no common data for the groups. This, in turn, makes it impossible to adjust for differences between the groups. Moreover, when "using the equivalent-groups design it is important that the groups be as similar as possible with respect to the ability being measured; otherwise, an unknown degree of bias will be introduced in the equating process" (Petersen et al., 1986, p. 245).

In practice, it is often impossible to randomly select the two groups to use in the test administration. In such a situation, the anchor-test design is used to adjust for random differences between the groups. In this design, "each test contains a set of common items, or a common external anchor-test is administered to the two groups simultaneously with the tests" (Hambleton & Swaminathan, 1985, p. 198). In theory, the anchor-test should consist of test items like those on the forms to be equated. Thus, the higher the correlation between the scores on the anchor test and the scores on the tests to be equated, the more useful the data from the anchor test will be. In other words, as Angoff (1971) has noted,

> If, for example, $r_{XU} = 0$ (and, presumably, $r_{YU} = 0$, since X and Y are parallel forms), this would indicate that observations made on Form U are irrelevant to the psychological functions measured by Form X or Form Y and are therefore not useful in making adjustments in those measures (p. 577)

where Form U is the anchor test. The length of the anchor test is suggested (Angoff, 1971) to be at least 20 items or 20% of the number of items in each test, whichever is larger.

Other methods of equating tests scores, which have been recently developed, include item response theory (Lord, 1980), confirmatory factor analysis (Rock, 1982), and section pre-equating (Holland & Wightman, 1982). Nonetheless, these methods require complex calculations. Moreover, since these methods are still being tested out in the field, "it is quite safe to assume that in the near future the more traditional methods will continue to play an important role in most testing programs" (Budescu, 1985, p. 14).

Linear Equating

Single-Group Design

As stated previously, select a large heterogeneous group. Divide this group into two random subgroups. Administer Form X to the first group and Form Y to the second group. According to Angoff (1982),

> Two scores, one on Form X and the other on Form Y- again, where X and Y are equally reliable and parallel measures- may be considered equivalent if their respective standard score deviates in any given group are equal. (p. 56-57)

In other words, two scores are equivalent if

$$\frac{(X - M_X)}{S_X} = \frac{(Y - M_Y)}{S_Y}$$

(Angoff, 1982). To solve for *Y*, first multiply both sides of the equation by S_Y. Thus,

$$\frac{(X - M_X)}{S_X} = \frac{(Y - M_Y)}{S_Y}$$

becomes

$$\frac{S_Y(X - M_X)}{S_X} = Y - M_Y$$

Adding M_Y to both sides,

$$\frac{S_Y(X - M_X)}{S_X} + M_Y = Y$$

Rearranging terms,

$$\frac{S_Y}{S_X}X - \frac{S_Y}{S_X}M_X + M_Y = Y$$

$$\frac{S_Y}{S_X}X + M_Y - \frac{S_Y}{S_X}M_X = Y$$

Letting A $= \frac{S_Y}{S_X}$, and B = $M_Y - \frac{S_Y}{S_X}M_X$,

$$Y = AX + B$$

where A is the slope and B is the intercept of the conversion equation.

Suppose that group one takes Form X and obtains a mean of 75 (i.e., $M_X = 75$) and a standard deviation of 8 (i.e., $S_X = 8$). Also, suppose that the mean and standard deviation for the second group are 70 and 9, respectively (i.e., $M_Y = 70$ and $S_Y = 9$). Substituting into

$$\frac{S_Y}{S_X}X + M_Y - \frac{S_Y}{S_X}M_X = Y$$

gives

$$\frac{9}{8}X + 70 - \frac{9}{8}(75) = Y$$

$$Y = \frac{9}{8}X + 70 - \frac{9}{8}(75)$$

Thus, a score of 80 on Form X is equivalent to

$$Y = \frac{9}{8}(80) + 70 - \frac{9}{8}(75)$$

$$Y = 75.625$$

a score of 75.625 on Form Y.

Random-Groups Design

As in the single-group design, select a large group and divide into two random subgroups. However, this time the original large group should be homogeneous. Once the group has been subdivided, administer Form X followed by Form Y to the first group. To the second group, administer the forms in the opposite order. That is, administer Form Y followed by Form X to the second group. According to Angoff (1982), in this design, "it is assumed that the standardized practice effect of Form X on Form Y is the same as the standardized practice effect of Form Y on Form X" (p. 59). Although the linear equation to be used in the random-group design is the same as the one for the single-group design ($Y = AX + B$), the A and B terms are given by different formulas. The linear equation for equating using random-group design is

$$Y = AX + B$$

where

$$A = \sqrt{\frac{S_{y_1}^2 + S_{y_2}^2}{S_{X_1}^2 + S_{X_2}^2}}$$

$$B = \frac{M_{Y_1} + M_{Y_2}}{2} - \frac{A(M_{X_1} + M_{X_2})}{2}$$

(Angoff, 1971).

Suppose that on a given test administration, group one obtains a mean of 60 and standard deviation of 5.5 on Form X. Similarly, when group one was administered Form Y, the group's mean and standard deviation were 55 and 5, respectively. Also, suppose that the second group's mean and standard deviation on Form X were 63 and 4.5, respectively. Likewise, suppose that the second group's mean on Form Y was 60 and that the group's standard deviation on Form Y was 5. Table 1 presents the different means and standard deviations for both groups. Substituting these values into

$$A = \sqrt{\frac{S_{y_1}^2 + S_{y_2}^2}{S_{X_1}^2 + S_{X_2}^2}}$$

$$B = \frac{M_{Y_1} + M_{Y_2}}{2} - \frac{A(M_{X_1} + M_{X_2})}{2}$$

yields

$$A = \sqrt{\frac{4.5^2 + 5^2}{5.5^2 + 5^2}} = .91$$

$$B = \frac{63 + 60}{2} - \frac{.91(60 + 55)}{2} = 9.175$$

So that,

$$Y = AX + B$$

$$= .91X + 9.175$$

Thus, a score of 70 on Form X is equivalent to

$$Y = .91X + B$$

$$= .91(70) + 9.175$$

$$= 72.175$$

Insert Table 1 About Here

Anchor-Test Design

Unlike the single-group design where the group takes both test forms or the random-group design where each group takes both test forms countered balanced, when using the anchor-test design each group takes each of the forms to be equated and a common (anchor) test.

Administer Form X to group one, Form Y to group two, and let U be the set of scores on the anchor-test. According to Crocker and Algina (1996), the assumptions in an anchor-test design equating are

1. The slope, intercept, and standard error of estimate for the regression of X on U in subpopulation 1 are equal to the slope, intercept, and standard error estimate for the regression of X on U in the total population.
2. The slope, intercept, and standard error of estimate for the regression of Y on U in subpopulation 2 are equal to the slope,

intercept, and standard error of estimate for the regression of Y on U in the population. (p. 460)

The linear equation for equating using the anchor-test design is, again,

$$Y = AX + B$$

where

$$A = \frac{S_Y}{S_X}$$

$$B = M_Y - AM_X$$

$$S_Y = \sqrt{S_{Y_2}^2 + b_{YU_2}^2 (S_U^2 - S_{U_2}^2)}$$

$$S_X = \sqrt{S_{X_1}^2 + b_{XU_1}^2 (S_U^2 - S_{U_1}^2)}$$

$$M_Y = M_{Y_2} + b_{YU_2} (M_U - M_{U_2})$$

$$M_X = M_{X_1} + b_{XU_1} (M_U - M_{U_1})$$

where b_{XU_1} is the slope of the regression of X on U in group one and b_{YU_2} is the slope of the regression of Y on U in group two (Angoff, 1971).

Suppose that both groups, one and two, have been administered their respective test forms. The groups' hypothetical data is presented in Table 2. Substituting the corresponding values yields

$$A = \sqrt{\frac{4.9^2 + .75(6.1^2 - 5.3^2)}{5.8^2 + 1.3^2(6.1^2 - 6.5^2)}} = 1.08$$

$$M_Y = 70 + .75(78 - 72) = 74.5$$

$$M_X = 73 + 1.3(78 - 80) = 70.4$$

$$B = 74.5 - 1.08(70.4) = -1.53$$

So that,

$$Y = AX + B$$

$$= 1.08X - 1.53.$$

Thus, a score of 70 on Form X is equivalent to

$$Y = AX + B$$

$$= 1.08X - 1.53$$

$$= 1.08(70) - 1.53$$

$$= 74.07$$

Insert Table 2 About Here

This paper has presented how to equate test scores using each of the following four designs: single-group, random-group, equivalent-group, and anchor-test. Each design has its own data collection assumptions to meet. Three of the four methods were illustrated by means of hypothetical situations. All four methods try to equate test scores from equally reliable and parallel measures. Although the anchor-test design is not as simple to implement as the other designs, it is one of the most popular equating procedures.

References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, DC: American Council of Education.

Angoff, W. H. (1982). Summary and derivation of equating methods used at ETS. In P. W. Holland & D. B. Rubin (Eds.), Test equating. New York: Academic Press.

Budescu, D. (1985). Efficiency of linear equating as a function of the length of the anchor test. Journal of Educational Measurement, 22, 13-20.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. San Diego, CA: Harcourt Brace Jovanovich College Publishers.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory principles and applications. Boston: Kluwer Nijhoff Publishing.

Holland, P. W., & Wightman, L. E. (1982). Section pre-equating: A preliminary investigation. In P. W.Holland & D. B. Rubin (Eds.), Test equating. New York: Academic Press.

Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. Educational Measurement: Issues & Practice, 7, 29-36.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), Educational measurement (3rd ed.). New York: Macmillan.

Rock, D. A. (1982). Equating using the confirmatory factor analysis model. In P. W. Holland & D. B. Rubin (Eds.), Test equating, New York: Academic Press.

Figure 1

Single-Group Design

	Form X	Form Y
Group A	*	*

Random-Groups Design

	Form X		Form Y	
	1st	2nd	1st	2nd
Group A	*			*
Group B		*	*	

Equivalent-Group Design

	Form X	Form Y
Group A	*	
Group B		*

Anchor-Test Design

	Form X	Form Y	Form U
Group A	*		*
Group B		*	*

Table 1

Means and Standard Deviations for Both Groups

	Mean		Standard Deviation	
	Form X	Form Y	Form X	Form Y
Group One	60	55	5.5	5
Group Two	63	4.5	60	5

Table 2

Hypothetical Data Anchor-Test Design

		Form X	Form Y	Form U
Group One	Mean	73		80
	SD	5.8		6.5
	b_{XU_1}	1.3		
Group Two	Mean		70	72
	SD		4.9	5.3
	b_{XU2}		.75	
Total	Mean			78
	SD			6.1

TM032247

U.S. Department of Education
Office of Educational Research and Improvement (OERI)

National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Equating Test Scores Using the Linear Method: A Primer	
Author(s): Jesus Tanguma	
Corporate Source: Texas A&M University	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed t Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	**Level 2A**	**Level 2B**
↑ [X]	↑ [ ]	↑ [ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche o

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Jesús Tanguma	Printed Name/Position/Title: Jesús Tanguma Grad Student	
Organization/Address: Texas A&M University Educational Psychology Dept. College Station, TX 77843	Telephone: 409-862-8089	Fax:
	E-mail Address: jot1609@labs.tamu.edu	Date: 1/7/00

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM: